REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

AD-A251 562

...imated to average 1 hour per response including the time for reviewing instructions, searching existing data sources, ...nd reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this ...his burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...he Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

...PORT DATE	3. REPORT TYPE AND DATES COVERED FINAL 1 APR 88 - 31 OCT 91
4. TITLE AND SUBTITLE "CASE BASED REASONING FOR TACTICAL PLANNING" (U)	5. FUNDING NUMBERS 61102F 6227/00
6. AUTHOR(S) Dr. Marc Goodman	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cognitive Systems, Inc. 234 Church Street New Haven, CT 06510	8. PERFORMING ORGANIZATION REPORT NUMBER AFOSR-TR- 92 0584
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SPONSORING/MONITORING AGENCY REPORT NUMBER F49620-88-C-0058
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; Distribution unlimited	12b. DISTRIBUTION CODE UL

13. ABSTRACT (Maximum 200 words)

Over the course of the three year contract on CBR, Cognitive has focused on designing and building a commerical-quality Case-Based Reasoning shell. They have addressed issues in the methodology for building case-based applications, appropriate and useful interfaces and functionality for representing, indexing, retrieving, and adapting cases, and have proven the usefulness of the technology by constructing specific applications and comparing their performance to corresponding applications built with other techniques.

14. SUBJECT TERMS			15. NUMBER OF PAGES 5
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

Marc Goodman
Cognitive Systems, Inc.
234 Church Street
New Haven, CT 06510

DARPA #F49620-88-C-0058 Final Report

Abstract

Over the course of the 3-year DARPA contract on CBR, Cognitive has focused on designing and building a commercial-quality Case-Based Reasoning shell. We have addressed issues in the methodology for building case-based applications, the scalability and maintainability of these applications, appropriate and useful interfaces and functionality for representing, indexing, retrieving, and adapting cases, and have proven the usefulness of the technology by constructing specific applications and comparing their performance to corresponding applications built with other techniques.

Introduction

Cognitive's focus on CBR has been the creation of a commercial-quality Case-Based Reasoning shell which could serve as the basis for rapid development of CBR applications. Pursuant to those ends, our goals have been to identify methodologies for CBR application construction, to develop a suite of facilities for representing, indexing, retrieving and adapting cases from large case libraries, to identify the extent to which domain knowledge can be easily obtained and used in application construction, and to address issues of scalability and assess application development time in comparison to other technologies, most notably in applications developed with Rule-Based systems.

Much of the technology in Cognitive's recently released ReMind™ CBR Shell was explored and developed within the context of this contract. It is a measure of our success along these lines that both Harmon Associates in their industry newsletter and Esther Dyson in her *Release 1.0* newsletter concluded that of the 6 or so commercial CBR shells available at the present time, ReMind™ is by far the most complete and powerful and should be the choice of anyone interested in doing serious research into CBR with a commercial shell.

An additional measure of our success has been the large number and variety of related contracts Cognitive has generated in the area of CBR. These contracts include governmental and commercial support for the reimplementation of the CBR Shell into C++ on a variety of different platforms, conversion to a generic interface package and a commercially supported database for greater portability, beta sites who are using the tool to explore CBR in their respective application areas, and Small Business Innovative Research contracts to extend the technology and broaden its applicability. Most importantly, this technology has led to the development of full-scale, commercially deployed applications, including the Prism Telex Classifier, which won an Innovative Applications award from the AAAI in 1990, and has been in continuous daily operation at Chase Manhattan bank since October of 1989.

First Year Work

Our initial conception of the CBR Shell was based on Nearest Neighbor retrieval. The envisioned methodology for building applications was to start with an initial library of cases, to go through these cases one by one and retrieve Nearest Neighbors, and based on these retrievals to adjust weights on features used for retrieval. Additional emphasis was placed on facilities to construct higher-level or more salient features algorithmically from primitive features immediately available on cases. The first version of the CBR Shell, presented at the First DARPA Workshop on CBR, embodied this philosophy.

The shell at that point had a collection of tools for dealing with cases composed of numeric, symbolic, textual and pictorial data, including a case-comparison screen which would allow a current situation to be compared with the top five closest cases, a hierarchy editor which could be used to define symbolic inheritance relationships between values, a formula editor which allowed new features to be derived from more primitive features, a facility for adjusting weights on individual fields, and a data entry facility for creating cases representing new situations with a form-based interface. All of these tools were packaged within a Menu-driven, Graphical User Interface. Case libraries were kept in a home-brew B-tree database, designed specifically for use with the CBR Shell.

This version of the CBR Shell was applied to the Land Warfare Database (LWDB) provided by Data Memory Systems, Inc. to develop a Battle Plan Advisor. During the development of this prototype, certain problems arose. Our notion of altering retrieval by adjusting feature weights proved inadequate, as changes made to facilitate a particular retrieval would destroy effects achieved with previous adjustments. Further, explanation facilities with Nearest Neighbor were severely limited, in that it was difficult to determine which features of a current situation should be altered in order to cause different cases to be retrieved. Finally, Nearest Neighbor is an exhaustive algorithm that took significant work to retrieve cases. For these reasons, it was necessary to develop a more sophisticated indexing and retrieval algorithm.

A significant change during the first year of the contract was the adoption of an inductive discrimination tree formation algorithm for case indexing. This algorithm allowed the system to automatically select features which were good discriminators for the purpose of accounting for variance in the classifications or outcomes of cases, and to build a binary discrimination tree of those salient features which could be used at retrieval time. The first impact of this approach was that retrieval became significantly faster, from O(mn) with nearest neighbor (where m is the number of fields per case and n is the number of cases) to O(log n). A more significant impact, in terms of the methodology used to build applications, was that instead of iteratively adjusting feature weights in nearest neighbor, the inductive algorithm could automatically adjust itself to an arbitrarily large set of cases with good results. Finally, explanation of retrieval could be presented as the specific set of features used for retrieval, and it was easily apparent which features could be altered in a particular situation to allow different cases to be retrieved.

Significant changes were made in the user interface to accommodate inductive clustering. A cluster editor was created which allowed the binary discrimination tree and the set of indexed cases to be displayed. Interfaces were also created to allow a knowledge engineer to optionally guide the clustering process at a variety of levels of interaction. Facilities were created to allow the knowledge engineer to create customized texts which could be displayed instead of the raw discrimination, so that an end-user could deal with English descriptions rather than statements of inequality. A form-based importance editor was created which allowed a knowledge engineer to designate features to be used in matching, as the outcome to be predicted, or as ignorable. The older feature weighting mechanism and interface were de-released. The case comparison interface was de-released in favor of a set of subtools which allowed a user to view factors used for retrieval, scan through a list of retrieved cases one at a time, and to compare a current situation to one of the retrieved cases in a form-based, side-by-side display. The Battle Plan Advisor was rebuilt in this version of the CBR Shell and received favorable attention at the 1989 meeting of the West Point Senior Officer AI Seminar and at the Army War College.

Second Year Work

Our work during the second year focused on enriching the tools for case representation and on incorporating user-provided domain knowledge to augment case representation and retrieval. During this period, an initial set of experiments were conducted comparing the efficacy of a Case-Based and an existing Rule-Based approach to telex classification. A new set of field types was created for the Shell, including sets of symbolic values. These field types required extensions to the data entry facilities, clustering tools and editors, and retrieval mechanism. The text of the telex was run through a lexical pattern matcher and the resulting set of tokens was saved as a set of symbols. The results were quite favorable for CBR; a system with equal accuracy was created in roughly 25% of the time needed to create the rule-based system. Further, the case-based system was significantly more maintainable and extendible than the rule-based system. This work was the basis for the Prism system, which won AAAI's Innovative Application award in 1990, and is the basis for ongoing work at Cognitive on CBR/Text, a case-based text classification and extraction tool with a variety of current government and corporate customers.

User-provided domain knowledge was directly applied to case indexing and retrieval in two ways. The first method was through the creation of "Case Prototypes," or conceptual categories of cases which reflected an expert's natural partitioning of situations in a domain. An interface was developed which allowed an expert to specify a set of features which defined a category and to create inheritance relationships between these categories. Cases were then automatically indexed underneath these prototypes (which were displayed in the cluster editor), and clustering proceeded on sets of cases within each category. In other words, the system now had a two-tiered memory organization with the top-tier formed of hand-crafted case prototypes, and the second-tier formed of inductively generated discrimination trees.

The second method for applying domain knowledge to case indexing was the creation of facilities for defining a Qualitative Model of a domain, specifying how features of a case interacted causally. This model was then used to guide the clustering process, by providing intermediate nodes in a causal graph whose values could be heuristically derived for particular cases, and by rejecting discriminations which contradicted the Qualitative Model. At this point, the symbol editor was extended to allow the creation of partial orders between symbolic values for use in the Qualitative Model. The result was not only a more expressive tool, but an overall system which allowed a quite powerful methodology for building applications. This new methodology was to create an initial qualitative model of a domain through a cursory examination of a small set of initial cases, to inductively index a large number of cases using this model, to search for sets of cases for which no causally valid explanation exists within the model to explain variance in outcome, to compare differing cases within that set, and to elaborate the qualitative model to account for this variance. The process repeats until the model is adequate to account for all variance. By adding to the model at each point, the system is able to guarantee that any cases previously discriminated will continue to be discriminated, and that the expressive power of the model is strictly increasing. This technique also proved quite effective for finding inconsistencies in large bodies of data, since cases that only differed on unimportant features but that differed in classification would end up in the same cluster.

Third Year Work

Third-year work continued in the same vein as in the second year, with the focus on increasing the tools expressive power and on using expert-supplied domain knowledge. Case-type fields which allowed cases to point to other cases in the same or different case libraries were used both as pointers to context (as in Apple's computer diagnostic and recovery application) and as temporal links for representing evolving situations (as in Parse-O-Matic). Frame-type fields were also included and used in Parse-O-Matic, a system which performs natural language processing in the domain of stock screening. As with Prism, Parse-O-Matic was compared to an existing system in the same domain, in this case the KNET parser developed by Bob Strong at Cognitive. Using Parse-O-Matic, equal robustness was achieved in less than half the time required with the KNET parser.

Parse-O-Matic was noteworthy for a few reasons. First, it represented a theory of action, based on past experience, which was similar to the dynamic theories of activity of Agree and Chapman, and the situated activity theories of Suchmann. Second, with between 40 and 50 thousand cases, it remains one of the largest CBR systems ever built. Finally, the high cost of inducing over frames required facilities for selectively pre-indexing based on causally related features. In all, 10 separate passes of heuristically guided induction were used in indexing Parse-O-Matic's case library, in a manner akin to Barletta and Mark's two-pass explanation-based indexing of cases.

The third year also saw a dramatic increase in beta sites, including application areas at American Express, NCS, Lockheed, Apple, the USGS, OSD, and BRL. These beta sites require significant extensions to and modifications of interfaces and facilities in the shell. Studying the needs of these beta sites contributed greatly to our design of an adaptation interface, designed and developed in the third year, which required extensions to the formula editor, the case viewer, the qualitative model editor, and the retriever, as well as the introduction of new interfaces to explain adaptation.